PCPs and HDX - Lecture 1

November 8, 2016

Introduction

In this course we will study probabilistically checkable proofs (PCPs), local testability, and high dimensional expansion. All of these have to do with connecting local observations and global phenomena. Local-to-global phenomena are all around. A lot of what we see is "local observations" and we often want to draw global conclusions from them. Some aspects of this are studied in the field of property testing where we think of a big object (like the internet) that is too large for us to be able to access it all, and we can only study it through looking at randomized local pieces. Another reason to care about local-to-global is because of the relation to approximation, and NP hardness of approximation.

1 Expansion

A graph $G = (V, E)$ is a collection of vertices V and undirected edges E. The (edge) expansion of a graph is a property that measures the edge boundary of a set $S \subset V$, which is the number of edges between S and $V - S$ (denoted $E(S, V - S)$), with respect to the size of S (or $V - S$ if it is smaller). A graph is an expander if every set has a boundary that is proportional to the size of the set or its complement:

$$\Phi(G) = \min_{\phi \neq S \subset V} \frac{|E(S, V - S)|}{\min(|S|, |V - S|)} = \min_{\phi \neq S \subset V} \frac{Pr_{uv \in E}[|\{uv\} \cap S| = 1]}{\min(\Pr[S], \Pr[V - S])} \cdot \frac{|E|}{|V|}.$$

An unconventional view of graph expansion is to think of the graph as a system of local equality checks where we have a Boolean function labeling the vertices and each edge $uv \in E$ checks that $f(u) = f(v)$. We think of these checks as trying to confirm that the function f is a costant, i.e. either $f = \mathbf{1}$ or $f = \mathbf{0}$. The boundary is the collection of unsatisfied edges.

Claim 1.1. *Let $G = (V, E)$ be an undirected graph. Then it holds that*

$$\Phi(G) = \inf_{f : V \to \{0,1\}, f \neq \mathbf{0}, \mathbf{1}} \frac{\mathrm{rej}(f)}{\mathrm{dist}(f, \{\mathbf{0}, \mathbf{1}\})} \cdot \frac{|E|}{|V|}.$$

where $\mathrm{rej}(f) = \Pr_{uv \in E}[f(u) \neq f(v)]$, *and* $\mathrm{dist}(f, \{\mathbf{0}, \mathbf{1}\})$ *the relative distance of f from the constant functions $\mathbf{0}, \mathbf{1}$, i.e.* $\min(\Pr_v[f(v) = 0], \Pr_v[f(v) = 1])$*.*

Proof. Every set S can be written as a function $f = \mathbf{1}_S$, and it is easy to check that $\mathrm{rej}(f) = |E(S, V - S)|/|E|$ and also that $\mathrm{dist}(f, \{\mathbf{0}, \mathbf{1}\}) = \min(|S|, |V - S|)/|V|$. □

So the expansion $\Phi(G)$ measures how good the local checks reflect the global property (of being a constant). Indeed,

- If f is a constant function, $f \in \{\mathbf{0}, \mathbf{1}\}$, then no edge "rejects", i.e. $\mathrm{rej}(f) = 0$.

- If $\mathrm{dist}(f, \{\mathbf{0}, \mathbf{1}\}) = \delta$, then at least $\frac{|V|}{|E|} \cdot \Phi(G) \cdot \delta$ fraction of the edges reject.

The property of being a constant function is not very interesting, as it consists of only two functions, $\mathbf{0}, \mathbf{1}$. We will be interested in much richer properties (in particular, all NP languages!) that have a similarly expanding collection of local tests.

2 PCPs

The PCP theorem says that every NP statement can be encoded in such a way as to allow local verification of the statement.

Theorem 2.1 (PCP Theorem). *For every language L in NP, there is $\varepsilon > 0$ and $q \in \mathbb{N}$ and a randomized verifier that reads an input x, and then in polynomial time computes a list of polynomially many local tests, such that each test accesses at most q bits in a proof string π and accepts or rejects, and such that*

- *Completeness: If $x \in L$ there is a proof such that all of the local tests accept, i.e.* $\mathrm{rej}(\pi) = 0$.
- *Soundness: If $x \notin L$ for every proof π at least ε fraction of the tests reject, i.e.* $\mathrm{rej}(\pi) \geq \varepsilon$.

where $\mathrm{rej}(\pi)$ *denotes the fraction of local tests that reject a proof π.*

It is easy to create, for any language in NP, a distribution over local tests that satisfies the above with $\varepsilon = 0$. I.e. in case the input is not in the language we will know that at least one of the local tests reject. The important extra given by the PCP theorem is that if $x \not\ni L$ every proof will be caught with probability *at least* $\varepsilon > 0$.

For example, let us consider 3-Coloring. Given a graph $G = (V, E)$, the statement "G is 3-colorable[1]" is an NP statement: given a 3-coloring $c : V \to \{1, 2, 3\}$ a natural way to check if it is a legal coloring is by going through each of the edges and checking that the endpoints have different colors. This can easily be made into a distribution over local tests: each edge gives one local test that accesses two locations in the proof $c : V \to \{1, 2, 3\}$, which can be easily encoded by at most 4 bits. This procedure satisfies

- Completeness: If the graph is 3 colorable then there is a proof (in this case, a coloring $c : V \to \{1, 2, 3\}$) such that $\mathrm{rej}(c) = 0$.
- Soundness: If the graph is not 3 colorable then *every* proof (in this case, every coloring $c : V \to \{1, 2, 3\}$) satisfies $\mathrm{rej}(c) > 0$.

Although this is a complete and sound proof system, it is not *robust*. There could be false proofs that are almost always accepted. The main thing that we are missing is a guarantee of a constant $\varepsilon > 0$ fraction of rejecting local tests in the soundness item. Note that each step of the verification looks at only two colors, but the entire verification looks at every single color in c, and is sensitive to every step.

How would we construct a PCP for 3-coloring? One naive idea is to use the "random edge verifier". This verifier expects a 3-coloring of the vertices as proof and does the following: select an edge $uv \in E$ at random, read $c(u), c(v)$, reject iff $c(u) = c(v)$. Let us denote by $\mathrm{rej}(G, c)$ the probability that the random edge checker rejects the coloring c,

$$\mathrm{rej}(G, c) = \Pr_{uv \in E}[c(u) = c(v)].$$

[1] i.e., there is a coloring of the vertices with 3 colors so that every edge has two differently colored endpoints

If G is 3 colorable, then there is a valid coloring $c : V \to \{1, 2, 3\}$ such that $\text{rej}(G, c) = 0$. Also, if G is not 3-colorable then every c is not a valid 3 coloring, and $\text{rej}(G, c) > 0$. But is it true that $\text{rej}(G, c) > \varepsilon$ for some $\varepsilon > 0$?

It is easy to come up with examples for graphs G where the answer is no: start with a 3 colorable graph on $n > 3$ vertices and add edges to it until it stops being 3 colorable. The resulting graph is not 3 colorable but the last valid 3 coloring c satisfies all but one edge, so $\text{rej}(G, c) = \frac{1}{|E|}$.

So for this graph the random edge verifier does not have a constant ratio between the the fraction of rejecting edges and the distance to being a 3-coloring. However, we can try to run a polynomial-time algorithm on our graph to make it more of an "expander".

Theorem 2.2. *There is a constant $\varepsilon_0 > 0$ and a polynomial time algorithm that inputs a graph $G = (V_G, E_G)$ and outputs a graph $H = (V_H, E_H)$ such that*

- *If G is 3-colorable then* $\text{rej}(H) = 0$
- *If G is not 3 colorable then* $\text{rej}(H) > \varepsilon_0$

This theorem implies (and is equivalent to) the PCP theorem. Indeed the verifier for a language L will start withan input x, use the reduction from L to 3-Coloring (such a reduction exists thanks to the fact that 3-Coloring is NP-complete) to compute a graph G such that $x \in L$ iff G is 3-colorable, then use the algorithm from the above theorem to compute H, and then expect as proof a 3 coloring of H. The verifier will choose a random edge in H and accept iff it is properly colored. So the PCP proof for $x \in L$ is a 3-coloring of H. If $x \in L$ there is a 3 coloring for H causing the verifier to always accept. If $x \notin L$ then every coloring for H will cause the verifier to accept with some constant probability.

Remark [hardness of approximation] We remark that the theorem can be interpreted as a *hardness of approximation* result: it shows that it is NP hard to *approximate* $\text{val}(H)$ for a given graph H. Indeed, if there was an algorithm that on input a graph H outputs a number $\beta \in (0, 1]$ that equals $\text{val}(H)$ up to a multiplicative $1 \pm \varepsilon$ factor, and if $\varepsilon < \varepsilon_0/2$, then this algorithm can be used to decide if G is 3 colorable or not. Since 3-Coloring is NP-hard, approximating val is NP-hard.

Theorem 2.2 has the flavor of increasing the expansion of G (in a sense that will be defined below), but that is not formally so. This is related to a notion of "PCPs of Proximity", and it is interesting whether such a poly-time transformation exists.

3 PCPs and expansion

Let us go back to the "naive" random edge checker for 3-Coloring and show that if G had a certain expansion property, then this checker would have constant soundness.

Let $COL(G) \subseteq \{1, 2, 3\}^V$ be the set of all valid 3-colorings of G. This set is either empty or not, but we can't tell by looking at G because it is an NP-hard question.

The "expansion" of the random edge checker with respect to the property $COL(G)$ is the ratio between the fraction of rejecting edges, $\text{rej}(G, c)$, and between its distance to being a proper coloring, $dist(c, COL(G))$ when minimized over all possible colorings c. In other words,

$$\gamma(G, COL(G)) = \min_{c \notin COL(G)} \frac{\text{rej}(G, c)}{\text{dist}(c, COL(G))}.$$

Suppose G is a graph for which $\gamma(G, COL(G)) \geq \varepsilon > 0$. Then the random edge checker works for G:

- If G is not 3 colorable, then for $c \in COL(G)$ clearly $rej(G, c) = 0$. In other words, there is a "proof" c that causes the checker to always accept.

- If G is not 3-colorable, then $COL(G) = \phi$ and for every coloring $dist(c, COL(G)) = dist(c, \phi) = 1$. Since $\gamma(G, COL(G)) \geq \varepsilon$ we know that $rej(G, c) \geq \varepsilon \cdot 1 = \varepsilon$, so the edge checker rejects every coloring with probability at least ε. In other words, for every "proof" c the edge checker rejects with probability at least ε.

The conclusion is that for such graphs G we don't need to apply the transformation from G to H. Soundness already holds as is.

We won't elaborate on this here but graphs G with $\gamma(G, COL(G)) \geq \varepsilon$ a stronger feature holds: even if the graph is 3-colorable, it is promised to reject colorings with probability that is proportional to their distance from proper colorings.

It would be nice to have an algorithm for increasing the expansion of a graph G with respect to $COL(G)$: i.e. of converting G to an expanding graph H while maintaining the feature that if $COL(G) = \phi$ then $COL(H) = \phi$. It might appear somewhat unintuitive that the algorithm, being polynomial-time, does this without knowing if the $COL(G)$ is empty or not.

Theorem 2.2 goes in this direction but a full such theorem is not known.